nature portfolio

https://doi.org/10.1038/s41563-021-00935-2

Supplementary information

Mid-infrared radiative emission from bright hot plasmons in graphene

In the format provided by the authors and unedited

Supplementary Information for
Mid-Infrared Radiative Emission from Bright Hot Plasmons in Graphene

Laura Kim[1,†], Seyoon Kim[1,2,†], Pankaj K. Jha[1], Victor W. Brar[1,2,3,*], Harry A. Atwater[1,3,*]

1. Thomas J. Watson of Applied Physics, California Institute of Technology, Pasadena, California 91125, USA
2. Department of Physics, University of Wisconsin-Madison, Madison, WI 53711, USA
3. Kavli Nanoscience Institute, California Institute of Technology, Pasadena, California 91125, USA

[*]V.W.B.: vbrar@wisc.edu

[*]H.A.A.: haa@caltech.edu

[†]These authors contributed equally to this work.

1. Optical properties and emissivity calculations

The optical properties of ITO were modeled with the Drude model with the following parameters: the high frequency permittivity ε_∞=4, the effective electron mass m^*=0.28m_e, the scattering time τ=5 fs, and the background carrier density n_e=2.3×10^{20} cm^{-3}. The departure from the room temperature optical properties of ITO above sufficiently high electronic temperature (which is the case under the pulsed laser excitation) was calculated. It was previously studied that the plasma frequency of ITO is temperature-dependent due to the non-parabolicity of the conduction band of ITO[S1]. By modelling the non-parabolic band structure as $\frac{\hbar^2 k^2}{2m} = E + \frac{E^2}{E_g}$, where $1/E_g$ represents the degree of non-parabolicity, the temperature-dependent plasma frequency takes the following form:

$$\omega_p(\mu,T)^2 = \frac{e^2}{3m\pi^2}\int_0^\infty dE\left[\frac{2m}{\hbar^2}\left(E+\frac{E^2}{E_g}\right)\right]^{3/2}\left(1+\frac{2E}{E_g}\right)^{-1}\left[-\frac{\partial f(\mu,T)}{\partial E}\right] \quad \text{(S1)}$$

The temperature-dependent chemical potential, μ, was obtained for a given n and T by solving the following equation:

$$n(\mu,T) = \frac{1}{\pi^2}\int_0^\infty dE\,\frac{m}{\hbar^2}\left(1+\frac{2E}{E_g}\right)\left[\frac{2m}{\hbar^2}\left(E+\frac{E^2}{E_g}\right)\right]^{1/2} f(\mu,T) \quad \text{(S2)}$$

The value of $1/E_g$=0.4 eV^{-1} was assumed[S1]. The temperature-dependent Drude permittivity, $\varepsilon\left(\omega,\omega_p(T)\right) = \varepsilon_\infty - \frac{\omega_p^2}{\omega^2+i\gamma\omega}$, was used to calculate temperature-dependent emissivity of ITO.

The optical properties for graphene were determined by the conductivity model based on a local limit random phase approximation[S2,S3]. The dynamic conductivity is defined as a function of frequency, temperature and chemical potential as follows:

$$\sigma(\omega) = \sigma_{\text{intra}}(\omega) + \sigma_{\text{inter}}(\omega) \tag{S3}$$

$$\sigma_{\text{intra}}(\omega) = \frac{i2e^2 k_B T}{\pi\hbar^2(\omega + i\tau^{-1})} \log\left(2\cosh\left(\frac{E_F}{2k_B T}\right)\right) \tag{S4}$$

$$\sigma_{\text{inter}}(\omega) = \frac{e^2}{4\hbar} G\left(\frac{\hbar\omega}{2}\right) + \frac{ie^2\omega}{\pi}\int_0^\infty \mathrm{d}E \, \frac{G(E) - G\left(\frac{\hbar\omega}{2}\right)}{(\hbar\omega)^2 - (2E)^2} \tag{S5}$$

where $(E) = \frac{\sinh\left(\frac{E}{k_B T}\right)}{\cosh\left(\frac{E}{k_B T}\right) + \cosh\left(\frac{E_F}{k_B T}\right)}$.

The optical properties of SiN_x in the mid-infrared are modeled by a classical Maxwell-Helmholtz-Drude dispersion model[S4,S5]. For the simulations in the visible and near-infrared spectral ranges, the refractive index of SiN_x was assumed to be 2.

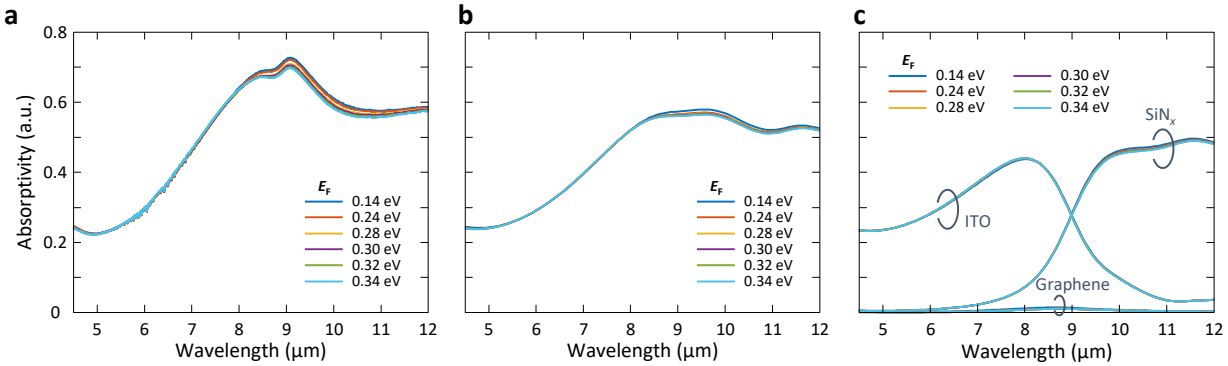

Figure S1. (a) Measured and **(b)** calculated Fermi-level-dependent absorptivity (emissivity) of the graphene/SiN_x/ITO device. (**c**) Calculated Fermi-level-dependent absorptivity (or emissivity) contribution from each layer at ambient temperature. These data were preprinted in Figs. 4.2 and 4.3 of ref. S6.

We used a finite element method to calculate the Fermi-level-dependent optical absorptivity of the device, consisting of a planar graphene layer on a 1 μm-thick SiN_x layer and a 50 nm-thick ITO layer, under plane wave incident using the optical properties described above. The calculated Fermi-level-dependent absorptivity (emissivity) of the device is in good agreement with the measured absorptivity at ambient temperature as shown in Figs. S1a and b. The emissivity contribution from each layer can also be extracted from the calculated electromagnetic power density. The results are shown in Fig. S1c. At the wavelength range longer than 9 μm, the majority of the emissivity comes from the SiN_x layer; at shorter wavelengths, that comes from the ITO layer. The majority of the Fermi level dependence comes from the SiN_x layer due to the Fabry-Perot mode formed in the SiN_x layer sandwiched between the graphene and ITO layers.

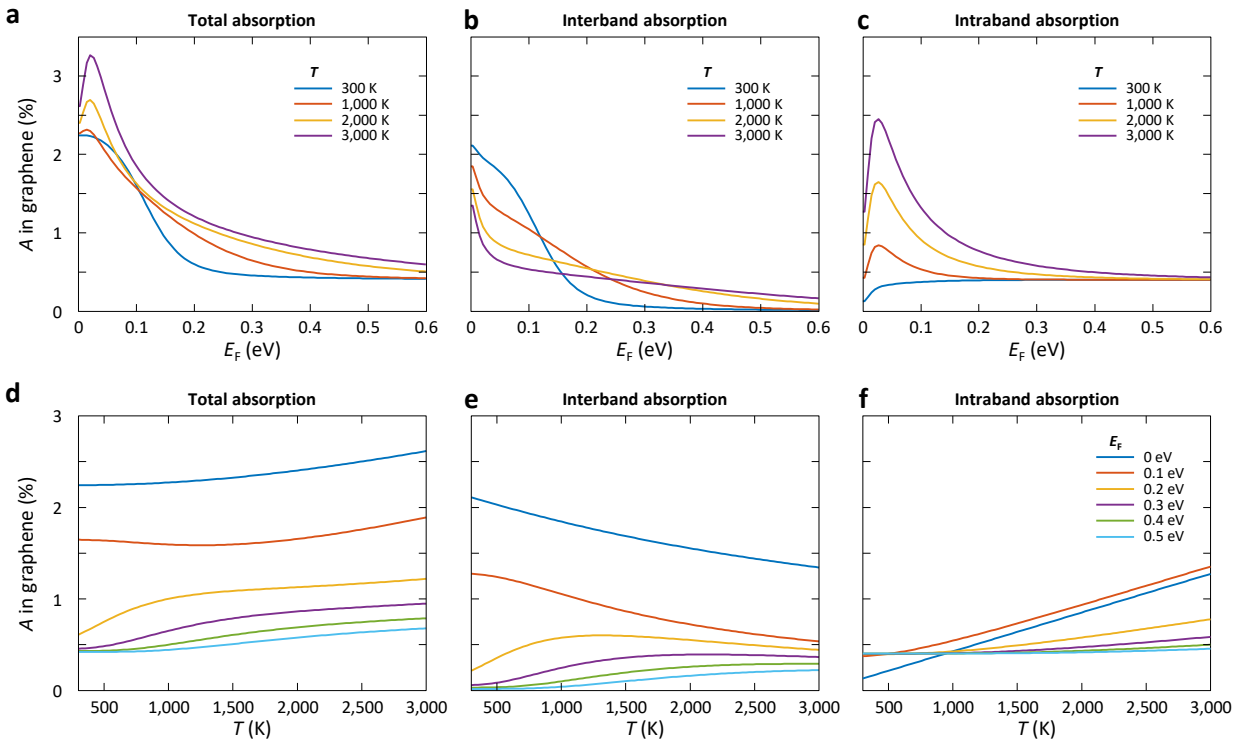

Figure S2. (**a**-**c**) Fermi-level-dependent and (**d**-**f**) carrier-temperature-dependent absorption and interband/intraband contributions of a suspended graphene (*i.e.*, structure: air-graphene-air) at the wavelength of 6 μm.

To understand temperature dependence of graphene emissivity, here we deconvolute and analyze interband and intraband contributions to the overall graphene emissivity/absorptivity. Figure S2b indicates that interband absorption is sharply suppressed for E_F > 0.2 eV at 300K. At an elevated carrier temperature, the interband contribution becomes non-negligible even for E_F > 0.2 eV due to the broadened Fermi distribution. However, we note that at any given temperature, the interband contribution decreases with increasing graphene Fermi level, which is an indication of Pauli-blocking (Figs. S2b and e). On the other hand, the intraband contribution increases with increasing E_F at 300 K, as expected. At an elevated carrier temperature (> 1,000 K), however, the intraband contribution peaks near E_F near CNP, as there exists a significant amount of thermally excited carriers (Fig. S2c). In Fig. S2f, there is a regime (< 1,000K) where the intraband contribution is slightly higher at a higher E_F; however, for E_F > 0.2 eV, temperature dependence of the intraband contribution is small, as shown in Fig. S2f. Therefore, as shown in Figs. S2a and d, the overall Fermi-level dependent trend over a wide temperature range from 300 K to 3,000 K is smaller of emissivity with an increasing graphene Fermi level.

In Figs. S3a-f, we plot the Fermi-level-dependent absorptivity (or emissivity) of graphene on a substrate (1 μm SiN_x-50 nm ITO) at various carrier temperatures. The overall spectral shape (*e.g.*, peak near 8 μm and dip near 11 μm) is dictated by the Fabry-Perot mode in the SiN_x membrane[S7,S8]. The Fermi-level-dependent trend for emissivity remains the same at various carrier temperatures.

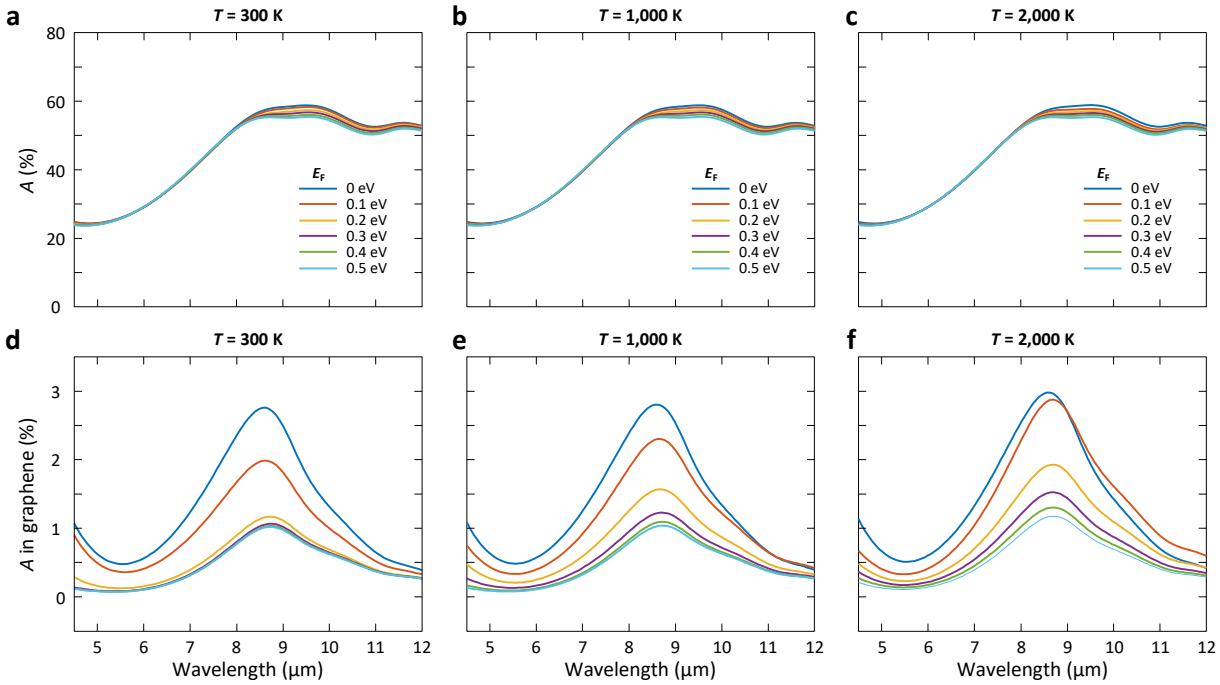

Figure S3. (**a-c**) Fermi-level-dependent absorption of Graphene/SiN_x/ITO with varying temperatures and (**d-e**) corresponding graphene absorption contributions.

2. Two-temperature model

A phenomenological two-temperature model for graphene was adopted from ref. S9. The electron and SCOP temperatures of graphene, T_e and T_p, were obtained by solving the following coupled equations:

$$\frac{\partial T_e}{\partial t} = \frac{I(t)}{\beta} - \frac{\Gamma_{e-p}}{C_e} \tag{S6}$$

$$\frac{\partial T_p}{\partial t} = \frac{\Gamma_{e-p}}{C_p} - \frac{T_p - T_o}{\tau_p} \tag{S7}$$

where $I(t)$ is the absorbed power density of the incident ultrafast pulse, $\Gamma_{e\text{-}p}$ is the electron-phonon energy exchange rate[S10,S11], C_e and C_p are the electronic and phonon specific heats, and τ_p is the lifetime of the strongly coupled optical phonons (SCOPs). It is assumed that all of the incident pulse energy initially goes to the electronic system, which loses its energy to the SCOPs with an exchange rate, $\Gamma_{e\text{-}p}$ (refs S9 and S12). The SCOP bath eventually achieve equilibrium with the rest of the phonons, whose temperature stays approximately at ambient temperature, T_o, with a characteristic time constant of τ_p. The temporal profile of the excitation optical pulse is modeled as $I(t) = \frac{A_{gra}F}{2\tau_{pulse}}\mathrm{sech}^2\left(\frac{t}{\tau_{pulse}}\right)$, where A_{gra} is the absorption in graphene at the wavelength of 850 nm (~2%), τ_{pulse} is the laser pulse duration (~100 fs), and F is the laser fluence. The electronic specific heat is analytically determined from the linear electronic band structure as $C_e(T_e) = \frac{18\xi(3)k_B^3T_e^2}{\pi(\hbar v_F)^2}$, where v_F is the Fermi velocity of graphene. The

specific heat of the SCOPs, C_p, and the electron-SCOP exchange rate, $\Gamma_{e\text{-}p}$, are expressed in ref. S9. τ_p is approximated as 1.5 ps (refs S9 and S12).

Electron-electron scattering occurs on the 10-fs timescale - much faster than other scattering mechanisms, such as phonon or photon emission processes. As a result, it is assumed that the photoexcited highly athermal electrons thermalize amongst themselves, and that they form a quasi-equilibrium state with a well-defined electronic temperature. At the start of this quasi-equilibrium state, it is assumed that photoexcitation energy is retained within the electronic system, and the corresponding chemical potential, μ, and carrier temperature, T_e, are obtained by simultaneously solving the equations in refs S9 and S10. Because the temperature and the specific heat are not well-defined before a thermalized carrier distribution is developed, the power density of the incident ultrafast pulse, $I(t)$, is scaled by a parameter, β, which has the same units as specific heat (eV m^{-2} K^{-1}) (ref. S10) to match the initial maximum electronic temperature determined from refs S9 and S10 that enforce carrier number and energy conservation.

$$N = N_p + N_o = \int f(E,\mu,T_e)D(E)\mathrm{d}E \tag{S8}$$

$$U = U_p + U_o = \int E \cdot f(E,\mu,T_e)D(E)\mathrm{d}E \tag{S9}$$

N_o and N_p are the initial Fermi-level-dependent 'dark' carrier density and photoexcited carrier density, respectively. The quantities, U_o and U_p, are the initial energy density and the energy density deposited by optical excitation, respectively. The photoexcited carrier density, N_p, was approximated as $N_p \approx \frac{A_{gra}U_p}{\hbar\omega_o}$, where A_{gra} is the graphene absorption at the wavelength of 850 nm, U_p corresponds to the pump fluence in J m^{-2}, ω_o is the laser frequency, $D(E)$ is the electron density of states in graphene, and $f(E,\mu,T_e) = \frac{1}{e^{(E-\mu)/k_B T_e}+1}$. The temperature profiles in graphene are Fermi-level-dependent; as the initial graphene Fermi level varied closer to the charge neutral point (CNP), the maximum electronic temperature increases.

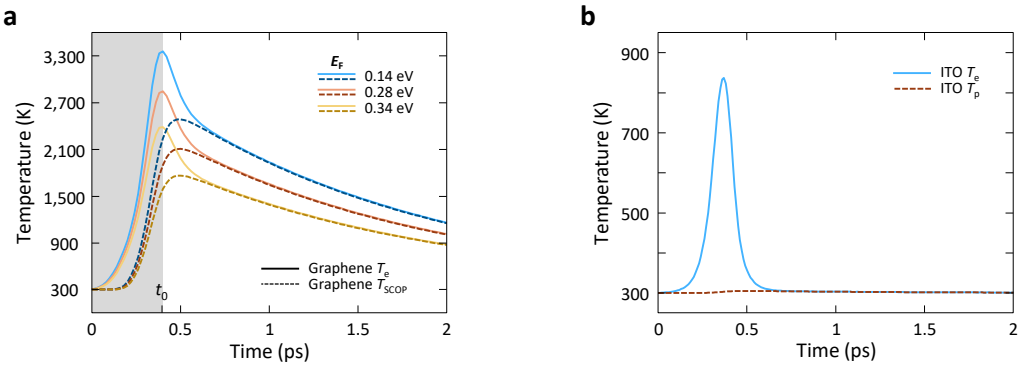

Figure S4. (a) Fermi-level-dependent temperature profiles of graphene and (**b**) ITO for a given laser fluence of 1.1 J m^{-2}. These data were preprinted in Fig. 3.2 of ref. S6.

The resulting Fermi-level-dependent temperature profiles for a given laser fluence are shown in Fig. S4a. The grey shaded region in Fig. S4a is when the system does not have a well-defined temperature as carriers have not yet reached a quasi-equilibrium state; this region is equivalent to stage (ii) in Fig. 1a of the main text.

The coupled partial differential equations for electron and phonon temperatures of ITO take the same forms as those for graphene as follows:

$$\frac{\partial T_{\mathrm{e}}}{\partial t} = \frac{I(t)}{\beta} - \frac{\gamma_{\mathrm{e-p}}}{C_{\mathrm{e}}} \tag{S10}$$

$$\frac{\partial T_{\mathrm{p}}}{\partial t} = \frac{\Gamma_{\mathrm{e-p}}}{C_{\mathrm{p}}} - \frac{T_{\mathrm{p}} - T_{\mathrm{e}}}{\tau_{\mathrm{p}}} \tag{S11}$$

The electron-phonon exchange rate is expressed in ref. S11, and it is a function of the Debye temperature, the Fermi velocity, the electron mean free path, and the Fermi energy. All of these parameters were deduced from the parameters used in the Drude model, specifically n_{e}, m^*, and τ. The values of v_{F}=7.84×10^5 m s^{-1}, l=3.92 nm, E_{F}=0.49 eV, and θ_{D}≈1,000 K were used. The temporal profile of the absorbed laser power density is approximated as $I(t) = \frac{A_{\mathrm{ITO}}F}{2\tau_{\mathrm{pulse}}d_{\mathrm{ITO}}}\mathrm{sech}^2\left(\frac{t}{\tau_{\mathrm{pulse}}}\right)$, where A_{ITO} is the absorption in ITO at the wavelength of 850 nm, d_{ITO} is the thickness of ITO, τ_{pulse} is the laser pulse duration, and F is the laser fluence. The electron heat capacity of ITO takes the analytical form for temperatures below the Fermi temperature as $C_{\mathrm{e}}(T_{\mathrm{e}}) = \frac{3\pi^2 n_{\mathrm{e}} k_{\mathrm{B}} T_{\mathrm{e}}}{\sqrt{36T_{\mathrm{F}}^2 + 4\pi^4 T_{\mathrm{e}}^2}}$, where n_{e} is the free electron density and T_{F} is the Fermi temperature. As the temperature dependence of the lattice heat capacity is small, C_{p} is assumed to be constant at 2.6×10^6 J m^{-3} K^{-1}. The value of β=28 J m^{-3} K^{-1} was chosen. The calculated temperature profiles of ITO are shown in Fig. S4b. The temporal profile of temperature of ITO is assumed to be graphene-Fermi-level-independent because the charge accumulation in the ITO layer due to the electrostatic gating is negligible for a given applied electric field over a 1 μm-thick SiN_x layer as described in the main text.

3. Emission from ITO-SiN_x-ITO

To show that the Fermi level dependence of the observe emission under pulsed laser excitation between 4.5 μm and 8.5 μm (as shown in Fig. 2a of the main text) originates from the Fermi level modulation in graphene, a control sample was fabricated and characterized. The control sample consists of a 1 μm-thick SiN_x substrate sandwiched between 50 nm-thick ITO layers, which served as top and bottom gates. The Fermi-level-dependent emission spectra from the control sample under pulsed laser excitation with a constant laser fluence of 1.12 J m^{-2} are shown in Fig. S5. The Fermi level dependence of the measured emission spectra from the graphene-less control sample is negligible.

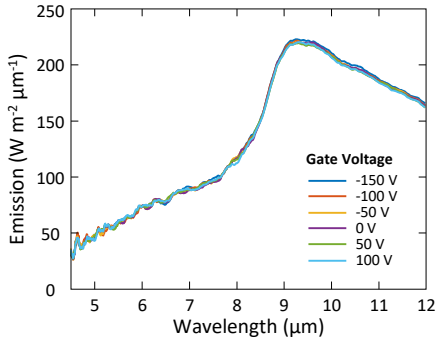

Figure S5. Measured emission spectra from ITO-SiN$_x$-ITO for various gate voltages under pulsed laser excitation with a constant laser fluence of 1.12 J m^{-2}. These data were preprinted in Fig. 4.6 of ref. S6.

4. Non-equilibrium plasmon dispersion and spontaneous emission spectra calculations

Page *et al*. reports a theoretical formulation to calculate exact complex-frequency plasmon dispersion with an arbitrary non-equilibrium carrier distribution[S13,S14]. The dielectric function, $\varepsilon(\omega,k_\mathrm{p})$, can be expressed within the random phase approximation as Eq. (S12).

$$\varepsilon\left(\omega,k_\mathrm{p}\right)=1-\frac{e^2}{2\varepsilon_0\varepsilon_\mathrm{eff}k_\mathrm{p}}\Pi_\mathrm{final}\left(\omega_\mathrm{p}+i\gamma_\mathrm{p},k_\mathrm{p}\right) \tag{S12}$$

where $\varepsilon_\mathrm{eff}=\frac{1+\varepsilon_{\mathrm{SiN}x}}{2}$ is the complex dielectric function of the substrate, SiN$_x$, and $\Pi_\mathrm{final}(\omega,k_\mathrm{p})$ is the dynamical polarizability. The graphene polarizability, Π, of an arbitrary non-equilibrium carrier distribution that is well-defined on the complex frequency plane is reported in ref. S15. The polarizability for the two-component plasma system with finite temperatures can be represented as the sum of the zero-temperature quasi-equilibrium polarizability and the correction terms that account for smearing of the Fermi edge due to finite temperatures as follows:

$$\Pi_\mathrm{final}=\Pi\left(\omega,k_\mathrm{p}\right)|_\mathrm{quasi-eq}^{T=0}+\int_0^\infty \mathrm{d}E\left[\frac{\partial\Pi|_{\mu=E}^{T=0}}{\partial E}\left(\delta f|_{\mu_\mathrm{c}}^{T}\right)+\frac{\partial\Pi|_{\mu=E}^{T=0}}{\partial E}\left(\delta f|_{\mu_\mathrm{v}}^{T}\right)\right] \tag{S13}$$

where $\Pi\left(\omega,k_\mathrm{p}\right)|_\mathrm{quasi-eq}^{T=0}$ is the zero-temperature quasi-equilibrium polarizability, μ_c and μ_v are the chemical potentials for conduction and valence bands, respectively, T is the shared temperature of the two-component plasma, and $\delta f=f(E)_\mu^T-f(E)_\mu^{T=0}$, where $f(E)=\frac{1}{e^{(E-\mu)/k_\mathrm{B}T}+1}$. The zero-temperature quasi-equilibrium polarizability is defined as follows:

$$\Pi\left(\omega,k_\mathrm{p}\right)|_\mathrm{quasi-eq}^{T=0}=\Pi\left(\omega,k_\mathrm{p}\right)|_{\mu_\mathrm{c}}^{T=0}+\Pi\left(\omega,k_\mathrm{p}\right)|_{\mu_\mathrm{v}}^{T=0}-\Pi\left(\omega,k_\mathrm{p}\right)|_{\mu=0}^{T=0} \tag{S14}$$

where

$$\Pi_\mu^{T=0}\left(\omega,k_\mathrm{p},\mu\right)=\Pi_+ + \Pi_-$$

$$= \frac{g}{8\pi\hbar^2 v_F^2}\left(-2\mu + \frac{k_p^2}{2\sqrt{k_p^2-\omega^2}}\left[\frac{2\mu-\omega}{k_p^2}\sqrt{k_p^2-(2\mu-\omega)^2} - i\operatorname{arccosh}\left(\frac{2\mu-\omega}{k_p}\right)\right]\right) +$$

$$\frac{g}{8\pi\hbar^2 v_F^2}\left(-2\mu + \frac{k_p^2}{2\sqrt{k_p^2-\omega^2}}\left[\frac{2\mu+\omega}{k_p^2}\sqrt{k_p^2-(2\mu+\omega)^2} - i\operatorname{arccosh}\left(\frac{2\mu+\omega}{k_p}\right)\right]\right) \quad \text{(S15)}$$

The finite-temperature correction terms in Eq. (S13) accounts for the Fermi edge smearing as quantified by $\delta f(E)|_\mu^T$, and the evaluation of the integrand of Eq. (S13) requires the derivative of $\Pi_\mu^{T=0}$, which can be written as follows:

$$\Pi'(\omega, k_p) = \frac{\partial \Pi_+}{\partial \mu} + \frac{\partial \Pi_-}{\partial \mu}$$

$$= \frac{g}{8\pi\hbar^2 v_F^2}\left(\frac{-2\sqrt{\mu-u}\sqrt{\mu-v}+2\mu}{\sqrt{u}\sqrt{v}} - 2\right) + \frac{g}{8\pi\hbar^2 v_F^2}\left(\frac{-2i\sqrt{i(\mu+u)}\sqrt{i(\mu+v)}-2\mu}{\sqrt{u}\sqrt{v}} - 2\right) \quad \text{(S16)}$$

where $u = \frac{\omega+k_p}{2}$ and $v = \frac{\omega-k_p}{2}$.

There are three parameters that define the final polarilzability, $\Pi_{final}(\omega_p+i\gamma_p,k_p)$: carrier temperature and chemical potentials for the particles and holes in the conduction and valence bands, respectively. We take a phenomenological approach to approximate time evolutions of these three parameters. For the temporal evolutions of carrier temperatures, the phenomenological two-temperature model is adopted (see Supplementary Information Section 2). For the temporal characteristics of carrier chemical potentials, we refer to the previous theoretical study on solving for non-equilibrium rate equations for an inverted massless Dirac fermion plasma in interaction with bosonic reservoirs via emission and absorption processes[S16]. The temporal evolution of chemical potential is phenomenologically fitted with three exponential functions with characteristic times: $\mu(t) = \mu_0 \sum_{i=1}^{3} A_i e^{-t/\tau_i}$. The solved complex-frequency plasmon dispersion relations for varying gate-controlled graphene Fermi levels and time are shown in Fig. S6. When solving for the complex-frequency dispersion, it was ensured that the solved dispersion curve does not cross a branch cut and remains continuous and physically meaningful. The details about relevant branch points/cuts and integration path are further explained in refs S13 and S14.

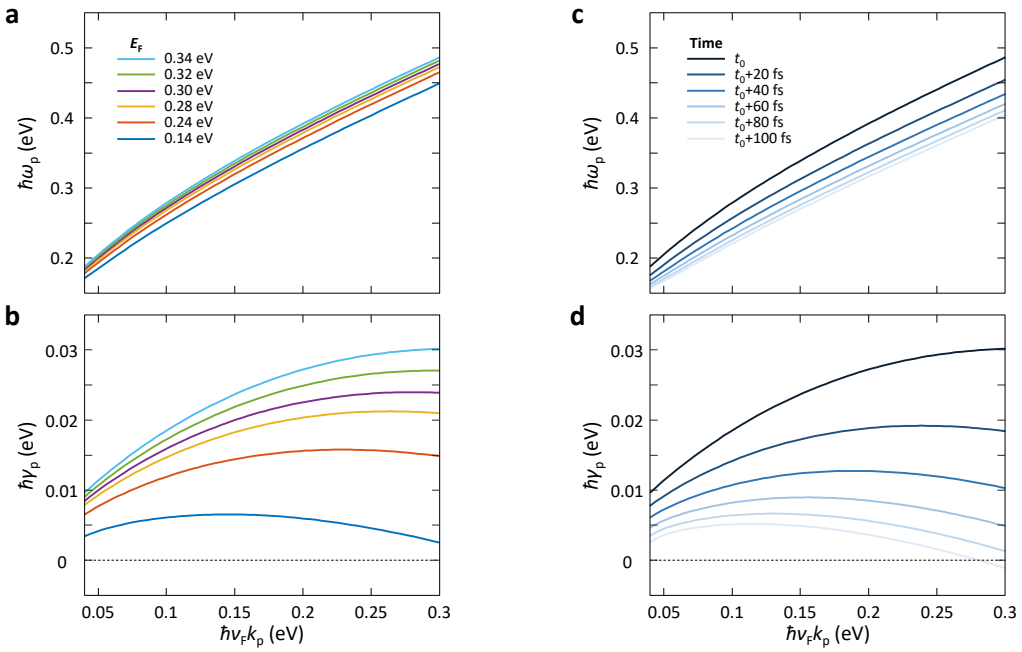

Figure S6. (**a**) The real part and (**b**) imaginary part of plasmon complex frequency dispersion for graphene on top of SiN_x with varying gate-controlled graphene Fermi levels at the time, t_0, at which a quasi-equilibrium is established. (**c**) The time-dependent real part and (**d**) imaginary part of plasmon complex frequency for a given initial graphene Fermi level of 0.34eV. These data were preprinted in Figs. 3.3 and 3.4 of ref. S6.

The spontaneous plasmon emission rate, g, is obtained by using Fermi's Golden rule in conjunction with the calculated complex plasmon dispersion relation. Under the first-order approximation, the plasmon emission rate takes the following semi-analytical form:

$$g \approx \alpha_f c k_\mathrm{p} \frac{\theta(\omega - v_\mathrm{F} k_\mathrm{p})}{\sqrt{\omega^2 - (v_\mathrm{F} k_\mathrm{p})^2}} \frac{2K(\omega, k_\mathrm{p})}{\frac{\partial \mathrm{Re}[\epsilon(\omega, k_\mathrm{p})]}{\partial \omega}}\Bigg|_{\omega = \omega_\mathrm{p}(k_\mathrm{p})} \tag{S17}$$

$$K(\omega, k_\mathrm{p}) = \int_{-1}^{+1} \mathrm{d}u \sqrt{1-u^2}\, f\left(\frac{\hbar(\omega + v_\mathrm{F} k_\mathrm{p} u)}{2}\right)\Bigg|_{\mu_\mathrm{c}}^{T} \times f\left(\frac{\hbar(\omega - v_\mathrm{F} k_\mathrm{p} u)}{2}\right)\Bigg|_{\mu_\mathrm{h}}^{T} \tag{S18}$$

The spontaneous emission spectra were calculated by scaling the plasmon emission rates by the plasmon density of states, $G = gD_\mathrm{p}$, where $D_\mathrm{p} = \frac{k_\mathrm{p}(\omega)}{2\pi} \frac{\mathrm{d}k_\mathrm{p}(\omega)}{\mathrm{d}\omega}$, assuming that the generated plasmons emit incoherently into all possible modes. The time-averaged spontaneous plasmon emission spectra for various gate-controlled graphene Fermi levels are shown in Fig. S7a. The increase in emission with increased hole doping of graphene can be intuitively understood as a result of enlarging the phase space for the excited carriers to relax by emitting plasmons. The time-averaged spontaneous emission spectra with increasing laser fluences are presented in Fig. S7b. As more carriers are added to the excited state population, more spontaneous plasmon emission is expected. The calculated plasmon emission spectra are scaled with the out-coupling efficiency discussed in the main manuscript (also in

Supplementary Information Section 6), and they are compared with the measured emission spectral under pulsed laser excitation in Fig. S7c.

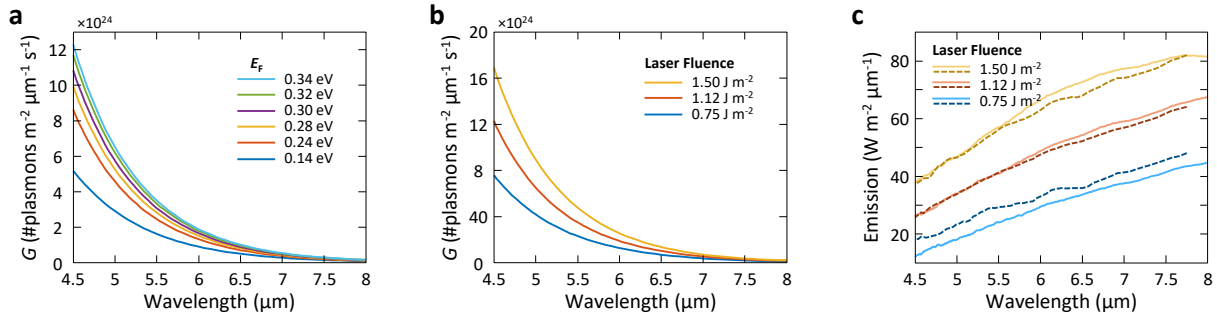

Figure S7. (**a**) Time-averaged spontaneous plasmon emission spectra for different gate-controlled graphene Fermi levels under a laser fluence of 1.12 J m^{-2}. (**b**) Time-averaged spontaneous plasmon emission spectra for a given gate-controlled graphene Fermi level of 0.34 eV under different laser fluences. (**c**) Laser-fluence-dependent measured emission spectra under pulsed laser excitation (solid color lines), compared with the calculated, out-coupled spontaneous plasmon emission spectra (dotted color lines). In both cases, the relevant gate-controlled graphene Fermi level is 0.34 eV. These data were preprinted in Figs. 4.10, 4.11, and 4.14 of ref. S6.

We also study the effects of collision losses. A collision time τ accounts for plasmon lifetime due to collisions with impurities, acoustic phonons, and optical phonons. It varies from tens to hundreds of femtoseconds depending on the quality of graphene and the underlying substrate. To include τ, the following transformation was performed on the polarizability $\Pi(\omega,k_\mathrm{p})$:

$$\Pi_\tau = \frac{(\omega+i\tau^{-1})\Pi\left(\omega+i\tau^{-1},k_\mathrm{p}\right)}{\omega+i\dfrac{\tau^{-1}\Pi\left(\omega+i\tau^{-1},k_\mathrm{p}\right)}{\Pi\left(0,k_\mathrm{p}\right)}}$$

$$= \frac{(\omega+i\tau^{-1})\left[\Pi_+\left(\omega+i\tau^{-1},k_\mathrm{p},\mu\right)+\Pi_-\left(\omega+i\tau^{-1},k_\mathrm{p},\mu\right)\right]}{\omega+i\dfrac{\tau^{-1}\left[\Pi_+\left(\omega+i\tau^{-1},k_\mathrm{p},\mu\right)+\Pi_-\left(\omega+i\tau^{-1},k_\mathrm{p},\mu\right)\right]}{\Pi\left(0,k_\mathrm{p},\mu\right)}} \qquad \text{(S19)}$$

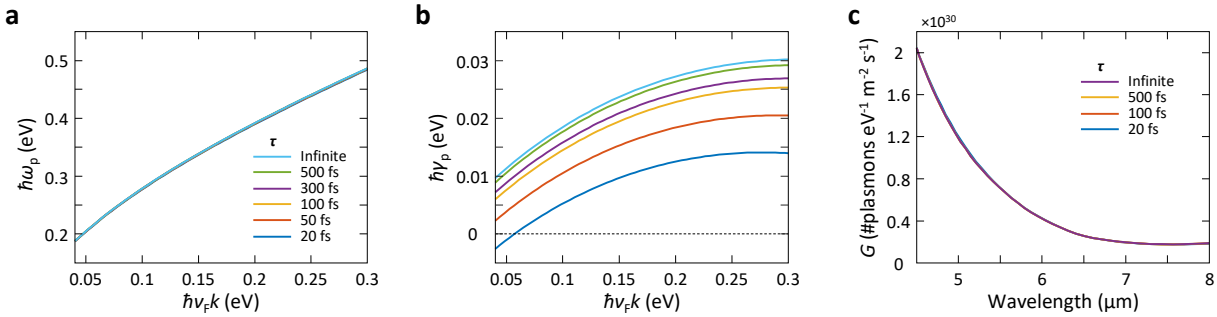

Figure S8. (**a**) The collision-loss-time (τ)-dependent real and (**b**) imaginary part of plasmon dispersion relation at t_0. (**c**) Calculated plasmon emission spectra with varying τ. The gate-controlled graphene Fermi level is assumed to be 0.34 eV. These data were preprinted in Figs. 3.5 and 3.6 of ref. S6.

The results are shown in Fig. S8. The scattering time, τ, has no significant impact on the real part of the plasmon frequency, while increasing the collision loss rate diminishes the imaginary part of the plasmon frequency, γ_p, suggesting decreasing net plasmon emission as additional absorption channels are introduced. The collision loss time, τ, does not explicitly enter the emission rate equation in Eq. (S17), and also it affects only weakly the plasmon dispersion relations[S13]. As a result, as shown in Fig. S8c, the calculated plasmon emission spectra show weak dependence on the collision loss rate.

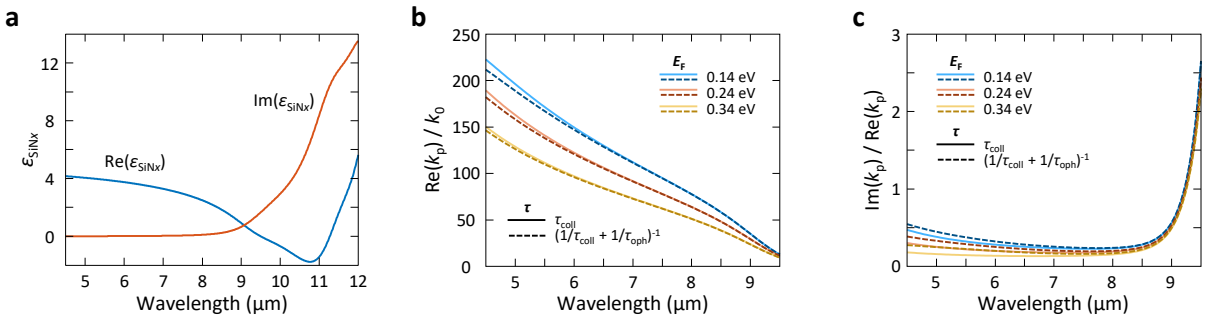

Figure S9. (**a**) Real and imaginary part of permittivity of the substrate, SiN_x. (**b**) Fermi-level-dependent graphene plasmon dispersion relations defined on a complex-k plane for a given carrier temperature of 2,000 K and total scattering loss time of $\tau=\tau_{coll}$ (solid lines) and $\tau=(1/\tau_{coll}+1/\tau_{oph})^{-1}$ (dashed lines), where τ_{coll}=50 fs. (**c**) Plasmon loss quantified by the ratio of imaginary and real parts of graphene plasmon wavevector, k_P. These data were preprinted in Figs. 3.7 and 3.8 of ref. S6.

Experimentally, the main source of the observed emission at wavelengths > 8 μm is thermal emission due to the substrate heating. The reason for this is that phonons of SiN_x strongly suppress plasmon excitation. As shown in Fig. S9a, the imaginary part of the substrate permittivity starts to rise at wavelength > 8 μm due to the substrate phonons. The complex-k plasmon dispersion relations were calculated for a given carrier temperature of 2,000 K and collision loss time of τ_{coll}=50 fs (Fig. S9b). In addition, the plasmon loss was quantified by calculating the ratio of the imaginary and real parts of the wavevector (Fig. S9c). At longer wavelengths (> 8 μm), the substrate phonons heavily damp graphene plasmon excitations as shown in Fig. S9c. In addition, Figs. S9a and c show that an additional damping channel originating from the optical phonons in graphene[S17,S18] does not have significant impact on the plasmon propagation behaviors and the resultant out-coupling efficiencies (Fig. S12d). The substrate phonon is the dominant damping channel for graphene plasmons.

5. Contributions of spontaneous and stimulated plasmon emission

There are three fundamental processes that contribute to the plasmon population in the system: stimulated emission, spontaneous emission and absorption (Fig. S10a). The ratio of stimulated to spontaneous emission rates is given by $\gamma_{stim}/\gamma_{spont}=n_p$, where n_p is the plasmon distribution function[S15]. The time evolution of n_p is obtained by

solving the rate equation, Eq. (S20), assuming that spontaneous emission, stimulated emission, absorption and collision loss are the processes changing the plasmon population[S15].

$$\frac{\mathrm{d}n_\mathrm{p}}{\mathrm{d}t} = \gamma_\mathrm{emiss}\left(n_\mathrm{p}+1\right) - \gamma_\mathrm{abs} n_\mathrm{p} - \gamma_\mathrm{coll} n_\mathrm{p} \tag{S20}$$

where γ_emiss and γ_abs are the plasmon emission and absorption rates, respectively, and γ_coll is the collision loss rate and is assumed to be constant. The plasmon distribution function, n_p, is solved for various values of $\gamma_\mathrm{coll}^{-1}$ ranging from 10 fs to 100 fs. In the frequency range of interest, which satisfies $\hbar\omega_\mathrm{p} < \mu_\mathrm{c}+\mu_\mathrm{v}$, plasmons experience interband gain while intraband processes are forbidden[S13,S14,S16]. Thus, when evaluating $\gamma_\mathrm{emiss/abs}$, only interband processes are considered. In first-order approximation, the plasmon interband emission/absorption rate[S13,S16], $\gamma_\mathrm{emiss/abs}$, is given by Eq. (S21)

$$\gamma_\mathrm{emiss/abs} \approx \frac{\alpha_f ck\theta\left(\omega - v_\mathrm{F}k_\mathrm{p}\right)}{\sqrt{\omega^2 - \left(v_\mathrm{F}k_\mathrm{p}\right)^2}} \frac{2K_\mathrm{emiss/abs}\left(\omega,k_\mathrm{p}\right)}{\frac{\partial \mathrm{Re}\left[\varepsilon\left(\omega,k_\mathrm{p}\right)\right]}{\partial\omega}}\Bigg|_{\omega=\omega_\mathrm{p}} \tag{S21a}$$

$$K_\mathrm{emiss}\left(\omega,k_\mathrm{p}\right) = \int_{-1}^{1} \mathrm{d}u\sqrt{1-u^2}\, f\left(\frac{\hbar\omega+\hbar v_\mathrm{F}k_\mathrm{p}u}{2},\mu_\mathrm{c},T_\mathrm{e}\right) f\left(\frac{\hbar\omega-\hbar v_\mathrm{F}k_\mathrm{p}u}{2},\mu_\mathrm{v},T_\mathrm{e}\right) \tag{S21b}$$

$$K_\mathrm{abs}\left(\omega,k_\mathrm{p}\right) = \int_{-1}^{1} \mathrm{d}u\sqrt{1-u^2}\left(1-f\left(\frac{\hbar\omega+\hbar v_\mathrm{F}k_\mathrm{p}u}{2},\mu_\mathrm{c},T_\mathrm{e}\right)\right)\left(1-f\left(\frac{\hbar\omega-\hbar v_\mathrm{F}k_\mathrm{p}u}{2},\mu_\mathrm{v},T_\mathrm{e}\right)\right) \tag{S21c}$$

Figure S10b shows the calculated plasmonic distribution function as a function of time for a given initial graphene Fermi level of E_F=0.34 eV at λ=6 μm, along with the corresponding inversion factor, defined as $f_\mathrm{inv} = \frac{2n_\mathrm{e}(t-t_0)}{n_\mathrm{e}(t_0)} - 1$, where $n_\mathrm{e}(t)$ is the instantaneous number density of excited carriers for a given frequency. When the plasmonic distribution function is greater than one (*i.e.*, above the dotted line in Fig. S10b), the stimulated emission dominates spontaneous emission. On a 100-fs timescale, stimulated emission dominates spontaneous emission for the collision time, τ_coll, longer than 10 fs. This opens path to coherent plasmon amplification on sub-100s fs time scale. Once the inversion is depleted, absorption starts to deplete the plasmon population rapidly and spontaneous emission starts to dominate stimulated emission. Experimentally, the cumulative emission on a longer time scale is measured as the laser pulse arrives at the sample every 12.5 ns. Figure S10c shows the time-averaged ratio of stimulated to spontaneous emission rates. The results show that the cumulative observed emission per excitation pulse (pulse every 12.5 ns) is mostly due to spontaneous plasmon emission.

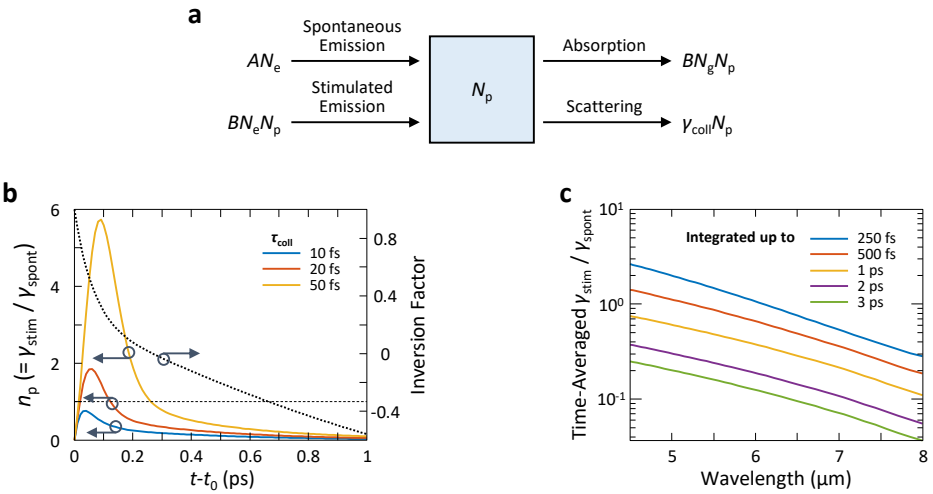

Figure S10. (**a**) Processes that contribute to the total number of plasmons in the system: spontaneous emission, stimulated emission, absorption, and scattering loss of plasmons. A and B are the spontaneous emission and stimulated emission/absorption Einstein coefficients, respectively. N_p is the plasmon density, and N_e and N_g are the carrier densities in the excited and ground states, respectively. γ_{coll} is the collision rate. (**b**) The calculated plasmonic distribution function (*i.e.*, ratio of stimulated to spontaneous plasmon emission rates) as a function of time for a given initial graphene Fermi level of E_F=0.34 eV at λ=6 μm, with varying collision loss rates. The corresponding inversion factor is plotted along the right y-axis in orange solid line. (**c**) Contributions of stimulated and spontaneous plasmon emission to the observed plasmon-coupled emission per excitation pulse. The time-averaged ratio for a given initial graphene Fermi level of E_F=0.34 eV and γ_{coll}^{-1}=20 fs as a function of wavelength when the ratio is integrated up to t-t_0=250 fs, 500 fs, 1 ps, 2 ps, and 3 ps. These data were preprinted in Figs. 3.9 and 3.10 of ref. S6.

6. Out-coupling efficiency of plasmons

The out-coupling efficiencies of plasmons are calculated based on the measured emission data. First, a thermal background expected for a given laser fluence is subtracted from the observed emission to determine the plasmon emission contribution. The out-coupling efficiencies of plasmons are determined by dividing the subtracted spectra by the calculated plasmon emission spectra. The results are shown as the dotted curves in Fig. S11a. The median of the dotted curves is shown in solid black curve in Fig. S11a. The out-coupling efficiency from a planar graphene is found to be in the order of 10^{-4}. The surface roughness, impurities, and fabrication residues are suspected to act as scattering centers.

To confirm that the spectral shape and the order of magnitude of the experimentally determined out-coupling efficiency, we performed two-dimensional full-wave simulations as depicted in Fig. S11b. The AFM measurement of a SiN_x surface showed the root mean square roughness of 0.4 nm as shown in Fig. S11c. The rough surface was modeled by $\sqrt{2}R_{\mathrm{RMS}}\cos(2\pi P_x x)$, where R_{RMS} is the root mean square surface roughness and the period of sinusoidal function P_x is assumed to be 6 nm. Graphene is modelled to conform to the roughness of the substrate, and graphene plasmons are launched within one period of the rough surface enclosed by perfect

electric conductor, which mimics a periodic boundary condition. In these simulations, we assumed the graphene carrier temperature of 2,000 K and the collision loss time τ_{coll} of 50 fs. The out-coupling efficiency was calculated by the radiation power toward air divided by the launching power of graphene plasmons. Here, a substrate consisting of a 1-μm SiN_x layer and a 50-nm ITO layer is assumed. Figure S11d confirms that the surface roughness is sufficient to scatter graphene plasmons into free-space. In particular, the spectral shape and the order of magnitude of the calculated out-coupling efficiency are consistent with the experimentally determined out-coupling efficiency. The difference between the measurement (Fig. S11a) and the calculation (Fig. S11d) is expected to originate from residues, foldings, and grain boundaries of graphene[S19] in a fabricated device that provide additional scattering centers but are difficult to address theoretically.

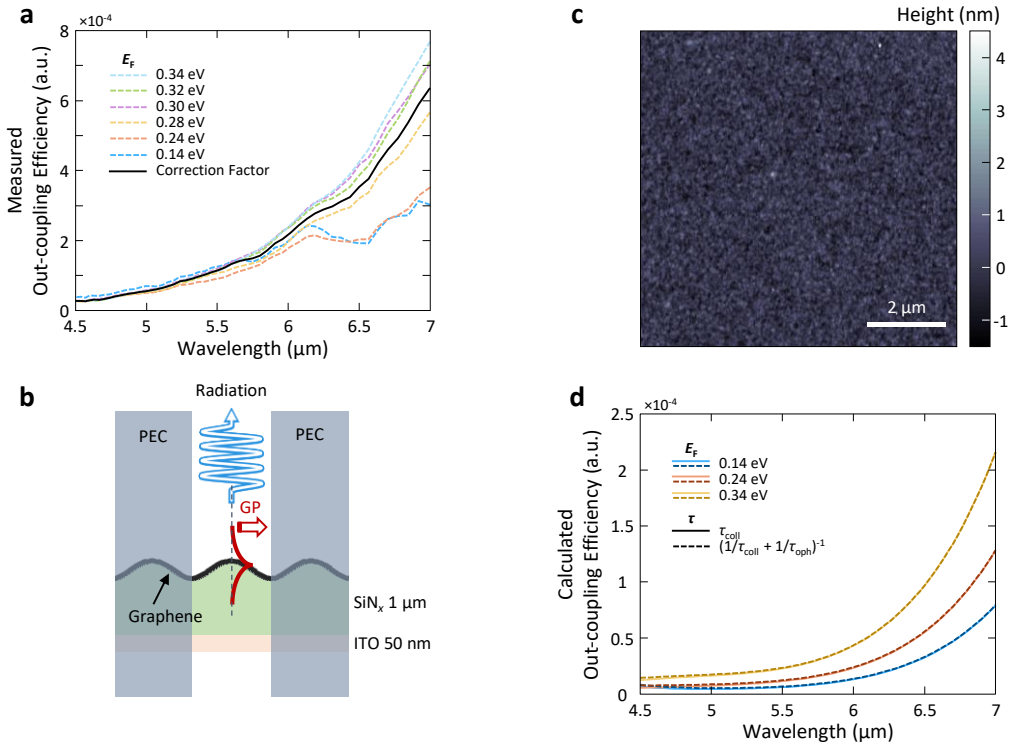

Figure S11. (**a**) Experimentally determined out-coupling efficiencies. (**b**) Schematic for calculating out-coupling efficiencies of graphene plasmons propagating on a planar graphene surface which conforms to the roughness of the underlying SiN_x. (**c**) AFM measurement of a SiN_x surface, showing root mean square roughness of 0.4 nm. (**d**) Calculated out-coupling efficiencies for a given carrier temperature of 2,000 K. Total scattering loss time is assumed as $\tau=\tau_{coll}$ for the solid lines and $\tau=(1/\tau_{coll}+1/\tau_{oph})^{-1}$ for the dashed lines, where τ_{coll}=50 fs. These data were preprinted in Fig. 4.12 of ref. S6.

7. Characteristics and effects of gold nanodisks (NDs)

The nanodisk (ND) arrays that are resonant and non-resonant with the incoming laser wavelength were made. The electric field intensity profiles of the resonant and non-resonant NDs on a semi-infinite SiN_x substrate are presented in Fig. S12a, and their optical cross-sections are shown in Fig. S12b.

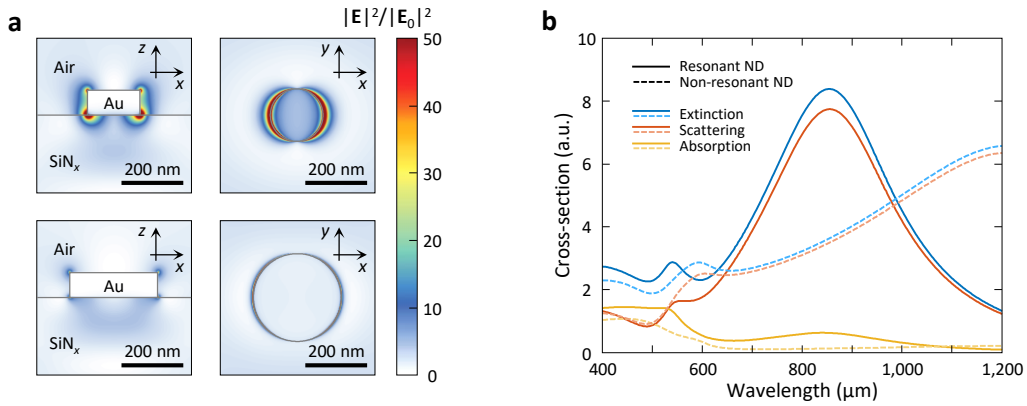

Figure S12. (**a**) Electric intensity distribution under the planar wave excitation at the laser wavelength of 850 nm. Side view electric intensity ($|\mathbf{E}|^2/|\mathbf{E}_0|^2$) distributions with the resonant (top left) and non-resonant (bottom left) NDs. Top view electric intensity ($|\mathbf{E}|^2/|\mathbf{E}_0|^2$) distributions for the resonant (top right) and non-resonant (bottom right) NDs. (**b**) Cross-sections of the resonant and non-resonant NDs. These data were preprinted in Figs. 4.15 and 4.16 of ref. S6.

The surface coverages of the resonant and non-resonant NDs are approximately 1% and 2.8%, respectively. They were randomly distributed over a graphene surface. Figure S13 compares the measured emissivity of the samples with and without the NDs. Due to the sparse distribution, the NDs have negligible effect on the overall emissivity of the sample in mid-infrared. In addition, full-wave simulation results suggest that the overall absorption as well as the absorption in graphene at the laser wavelength of 850 nm are not significantly altered by the NDs, as shown in Figs. S14a and b.

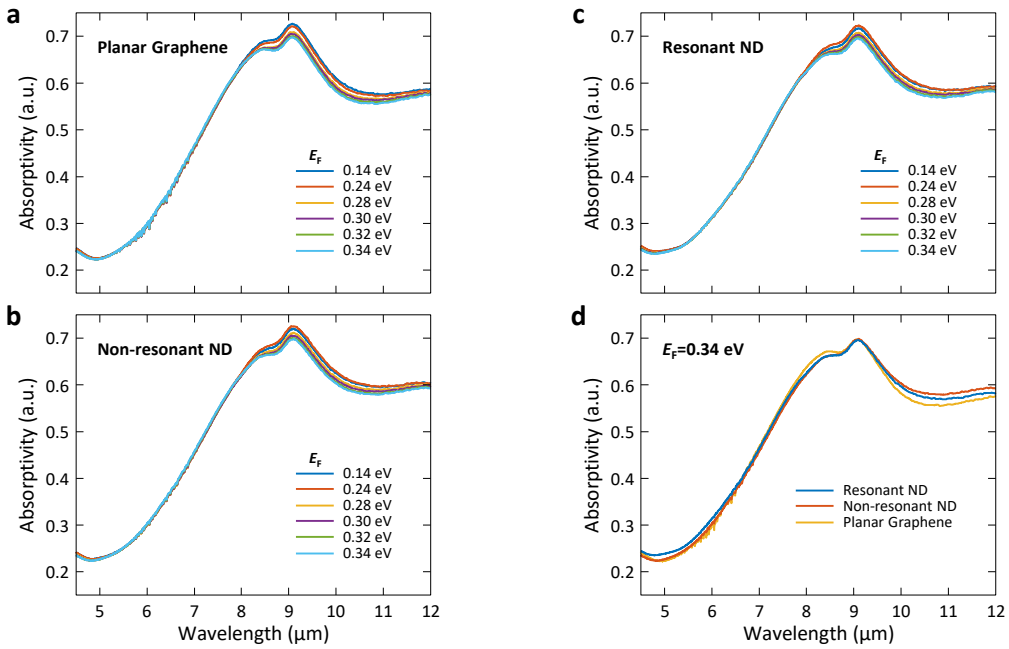

Figure S13. Fermi-level-dependent measured absorptivity (or emissivity) of the graphene/SiN_x/ITO structure at ambient temperature for (**a**) planar graphene (**b**) non-resonant nanodisks (NDs) on graphene, and (**c**) resonant nanodisks (NDs) on

graphene. (**d**) For a fixed graphene Fermi level at 0.34 eV, the measured emissivities with and without the NDs. These data were preprinted in Fig. 4.17 of ref. S6.

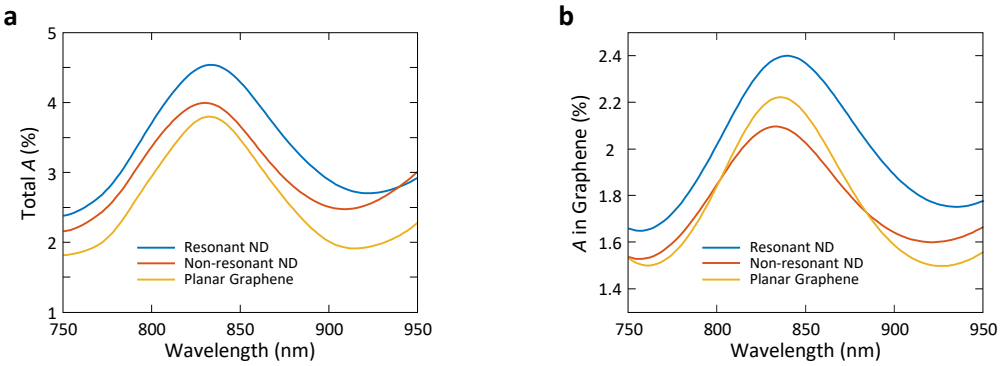

Figure S14. (**a**) Total absorptions of the graphene/SiN_x/ITO device with and without the NDs. (**b**) Absorptions in graphene with and without the NDs. These data were preprinted in Fig. 4.18 of ref. S6.

Figure S15a shows the measured emission spectra under pulsed laser excitation with a given fluence from planar and ND-decorated graphene samples. At shorter wavelengths (< 8 μm), we start to see deviations from the measured thermal emission profiles and increasing Fermi level dependence that are notably different from that seen in Fermi-level-dependent thermal emission as well as emission under CW laser excitation as shown in Figs. S15b and c, respectively. Such deviation and its Fermi level dependence are larger when graphene is decorated with the gold NDs for a given laser excitation fluence. Considering such a sparse distribution, however, the enhancement factor in the observed emission with the NDs cannot be properly described assuming their only role is to act as scatterers. In particular, the observed deviation and its Fermi level dependence is larger with resonant NDs than with non-resonant NDs; thus, not only geometric factors, but also electromagnetic factors contribute to such a large emission enhancement.

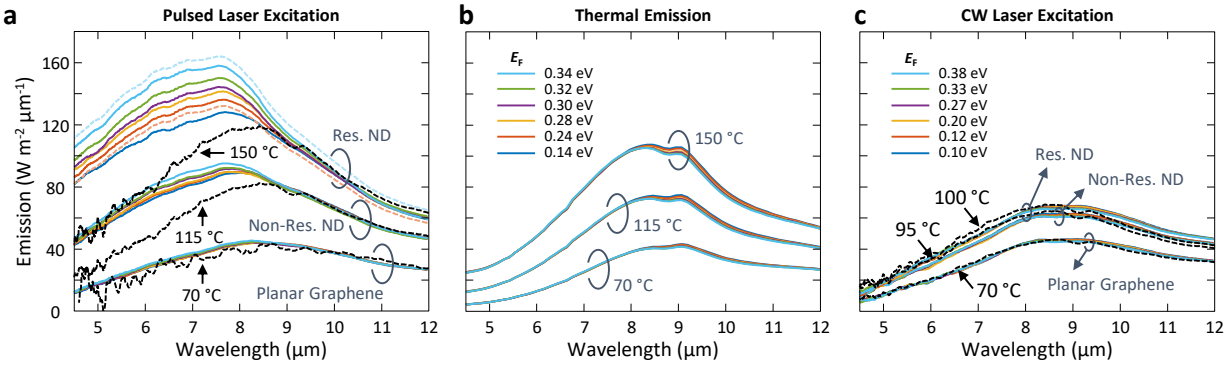

Figure S15. (**a**) Fermi-level-dependent emission spectra from planar and ND-decorated graphene samples under pulsed laser excitation with a laser fluence of 0.75 J m^{-2} (color solid lines). Each color corresponds to the graphene Fermi level denoted in the legend of Fig. S16b. In the resonant ND data, the dotted lines for 0.24 eV and 0.34 eV correspond to the original measurement, and the solid lines are fitted to compensate the offset, which appeared due to experimental

imperfections. Measured thermal emission spectra from the device for given temperatures of 70 °C, 115 °C, and 150 °C (black dotted lines). (**b**) Expected Fermi-level-dependent thermal emission spectra under isothermal conditions of given temperatures of 70 °C, 115 °C, and 150 °C. (**c**) Fermi-level-dependent emission spectra from planar and ND-decorated graphene samples under CW laser excitation with a laser fluence of 0.75 J m^{-2} (color solid lines). Measured thermal emission spectra from the device for given temperatures of 70 °C, 95 °C, and 100 °C (black dotted lines). These data were preprinted in Figs. 4.19, 4.20, and 4.21 of ref. S6.

The calculated field enhancement factor as a function of distance from a resonant/non-resonant ND is shown in Fig. S16b. The enhancement factor is more than 100 and 10 in the vicinity of a resonant and non-resonant ND, respectively. Although the NDs do not alter the overall absorption of the device at the laser wavelength due to the sparse distribution, the plasmonic resonance excited in the NDs can enhance absorption locally, creating more excited carriers. This effect will impact both population inversion and plasmon emission rates. Also, the field energy concentration at the graphene sheet greatly increases the coupling of the plasmons to the particle/hole plasma, enhancing the stimulated plasmon emission[S20].

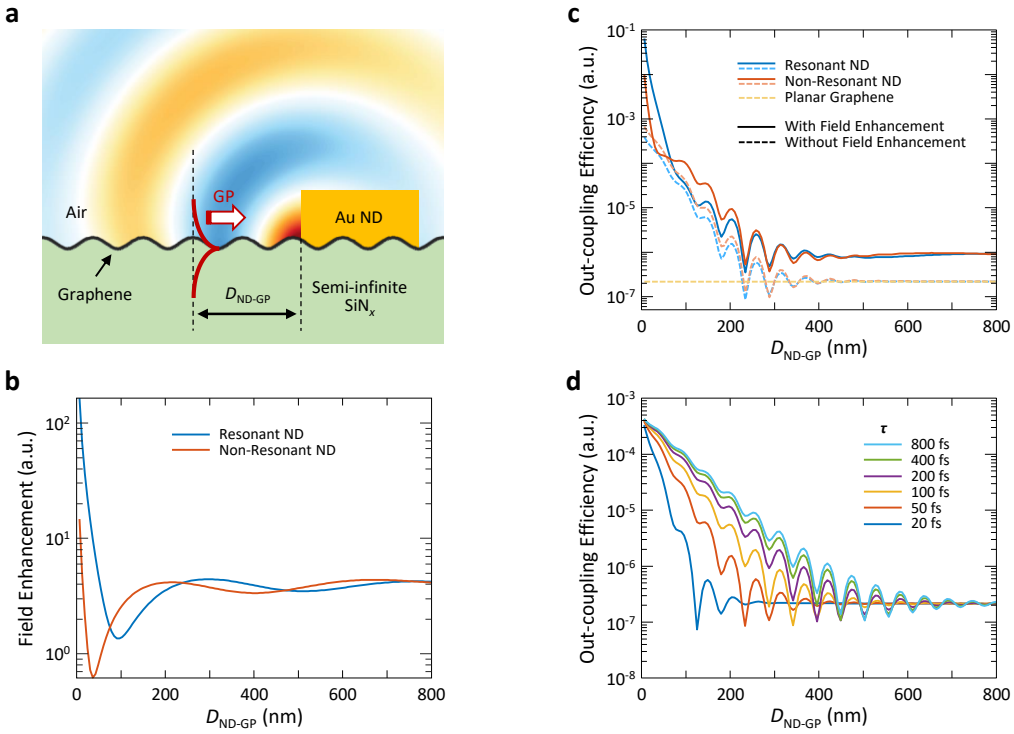

Figure S16. (**a**) Simulation schematic for calculating out-coupling efficiency of a single graphene plasmon launched at D_{ND-GP} distance away from a ND. Here, D_{ND-GP} denotes the distance between the position at which a graphene plasmon is launched and the edge of a ND that the launched graphene plasmon scatters off. A semi-infinite SiN_x substrate is assumed. (**b**) Field enhancement factor around the NDs due to the plasmonic resonance obtained from the transverse electric field profiles ($(|E_x|^2+|E_y|^2)/|\mathbf{E}_0|^2$) in Fig. S12a. (**c**) Calculated out-coupling efficiency of emitted plasmons with and without NDs with varying distance from the edge of a ND (D_{ND-GP}). The dotted lines represent the plasmon out-coupling efficiencies with resonant (blue) and non-resonant (red) NDs without considering the field enhancement (*i.e.*, considering purely geometric factors). The solid lines represent the plasmon out-coupling efficiencies with NDs considering the effects of the near-field enhancement. The yellow dotted line represents the out-coupling efficiency without NDs. All calculations

were performed at the wavelength of 6 μm for a given graphene Fermi level of 0.34 eV and the collision time (τ) of 50 fs. (**d**) τ-dependent out-coupling efficiency of a single graphene plasmon at the wavelength λ=6 μm launched with varying $D_{\text{ND-GP}}$. The carrier temperature of 2,000 K and the gate-controlled graphene Fermi level of 0.34 eV are assumed, and the field enhancement factors are not included. These data were preprinted in Figs. 4.23 and 4.24 of ref. S6.

We study the electromagnetic effects of the NDs on the out-coupling efficiency of graphene plasmons with two-dimensional full-wave simulations. The schematic for the simulations is shown in Fig. S16a, and the calculation results are presented in Fig. S16c. The dotted lines Fig. S16c are out-coupling efficiencies of a single graphene plasmon launched on a rough graphene surface in the presence of NDs without taking the near-field enhancement into consideration (*i.e.*, only considering geometric factors). The ripple patterns in the calculated out-coupling efficiencies in Fig. S16c are due to the interference between the scattered graphene plasmons from the nanodisk and those from the rough surface. If we assume they merely act as scatterers (considering only geometric factors), the out-coupling efficiency is slightly higher with the non-resonant NDs because its diameter is larger than that of the resonant NDs, serving as a more efficient out-coupling scatterer. When the field enhancement factor presented in Fig. S16b is taken into consideration, the calculated out-coupling efficiency in the presence of resonant NDs is approximately 10^5 times higher near the NDs compared to that on a rough surface of planar graphene. This suggests that the near-field enhancement takes part in improving out-coupling efficiency of graphene plasmons.

The out-coupling efficiencies are also calculated as a function of collision loss time, and the results are shown in Fig. S16d. The out-coupling efficiency of graphene plasmons increases as they are generated near a scattering site. For a given $D_{\text{ND-GP}}$= 250 nm, the out-coupling efficiency of graphene plasmons with τ=800 fs is approximately twenty times higher than that with τ=50 fs. This analysis shows that by carefully positioning scattering centers, we can greatly enhance the out-coupling and far-field collection efficiency of graphene plasmons. For example, an array of incorporated nanostructures separated by a distance that is comparable to the propagation length of plasmons will be able to efficiently out-couple plasmons before being damped.

8. Estimation of radiated output power of hot-plasmon-assisted mid-infrared radiation

We estimate the radiated average power of out-coupled plasmon emission of a feasible device structure by Eq. (S22)

$$P_{\text{out}} = Brightness \times A \times \Delta\lambda \times \int_0^{2\pi} \int_0^{\theta} \sin\theta \, d\theta d\phi \times \eta_{\text{out-coupling}} \times DC \qquad \text{(S22)}$$

where *Brightness* is the luminous intensity of plasmon emission in W m^{-2} μm^{-1} sr^{-1}, A is the device collection area, $\Delta\lambda$ is the bandwidth of plasmon emission, θ is the collection angle, $\eta_{\text{out-coupling}}$ is the out-coupling efficiency, and DC is the duty cycle of emission (limited by the laser duty cycle). The calculations, presented in Fig. 5 of the

manuscript and in Supplementary Information Section 7, suggest that an array of resonant gold nanodisks enhances the out-coupling efficiency approximately by three to four orders of magnitude compared to flat graphene, as shown in Fig. S16c. With a conservative approximation of the out-coupling efficiency of 10%, for a given pump fluence of 1.12 J m^{-2}, A=50×50 μm^2 and NA=0.58 of the Cassegrain lens used in this work, we expect approximately 61 μW output power for 80 MHz and 7.6 mW for DC = 10 GHz.

Supplementary References

S1. Guo, P. J., Schaller, R. D., Ketterson, J. B. & Chang, R. P. H. Ultrafast switching of tunable infrared plasmons in indium tin oxide nanorod arrays with large absolute amplitude. *Nat. Photonics* **10**, 267-273 (2016).

S2. Hanson, G. W. Dyadic Green's functions and guided surface waves for a surface conductivity model of graphene. *J. Appl. Phys.* **103** (2008).

S3. Falkovsky, L. A. Optical properties of graphene. *J. Phys. Conf. Ser.* **129**, 012004 (2008).

S4. Philipp, H. R. Optical properties of silicon-nitride. *J. Electrochem. Soc.* **120**, 295-300 (1973).

S5. Cataldo, G. *et al.* Infrared dielectric properties of low-stress silicon nitride. *Opt. Lett.* **37**, 4200-4202 (2012).

S6. Kim, L. Novel light emitting mechanisms originating from graphene plasmons near and far from equilibrium (California Institute of Technology, Pasadena, 2019).

S7. Kim, S. *et al.* Electronically tunable extraordinary optical transmission in graphene plasmonic ribbons coupled to subwavelength metallic slit arrays. *Nat. Commun.* **7**, 12323 (2016).

S8. Jang, M. S. *et al.* Tunable large resonant absorption in a midinfrared graphene Salisbury screen. *Phys. Rev. B* **90**, 165409 (2014).

S9. Lui, C. H., Mak, K. F., Shan, J. & Heinz, T. F. Ultrafast photoluminescence from graphene. *Phys. Rev. Lett.* **105**, 127404 (2010).

S10. Johannsen, J. C. *et al.* Direct view of hot carrier dynamics in graphene. *Phys. Rev. Lett.* **111**, 027403 (2013).

S11. Alam, M. Z., De Leon, I. & Boyd, R. W. Large optical nonlinearity of indium tin oxide in its epsilon-near-zero region. *Science* **352**, 795-797 (2016).

S12. Song, D. H. *et al.* Direct measurement of the lifetime of optical phonons in single-walled carbon nanotubes. *Phys. Rev. Lett.* **100**, 225503 (2008).

S13. Page, A. F., Ballout, F., Hess, O. & Hamm, J. M. Nonequilibrium plasmons with gain in graphene. *Phys. Rev. B* **91**, 075404 (2015).

S14. Page, A. F., Hamm, J. M. & Hess, O. Polarization and plasmons in hot photoexcited graphene. *Phys. Rev. B* **97**, 045428 (2018).

S15. Archambault, A., Marquier, F., Greffet, J. J. & Arnold, C. Quantum theory of spontaneous and stimulated emission of surface plasmons. *Phys. Rev. B* **82**, 035411 (2010).

S16. Hamm, J. M., Page, A. F., Bravo-Abad, J., Garcia-Vidal, F. J. & Hess, O. Nonequilibrium plasmon emission drives ultrafast carrier relaxation dynamics in photoexcited graphene. *Phys. Rev. B* **93**, 041408 (2016).

S17. Park, C. H., Giustino, F., Cohen, M. L. & Louie, S. G. Velocity renormalization and carrier lifetime in graphene from the electron-phonon interaction. *Phys. Rev. Lett.* **99** (2007).

S18. Yan, H. G. *et al.* Damping pathways of mid-infrared plasmons in graphene nanostructures. *Nat. Photonics* **7**, 394-399 (2013).

S19. Fei, Z. *et al.* Electronic and plasmonic phenomena at graphene grain boundaries. *Nat. Nanotechnol.* **8**, 821-825 (2013).

S20. Dubinov, A. A., Aleshkin, V. Y., Mitin, V., Otsuji, T. & Ryzhii, V. Terahertz surface plasmons in optically pumped graphene structures. *J. Phys.: Condens. Matter* **23**, 145302 (2011).